MODULE 2 : Estimation par intervalle de confiance

Il s'agit dans ce module de trouver une estimation par intervalle de confiance d'un paramètre θ, c'est-à-dire de construire une « fourchette de valeurs numériques permettant de situer » θ avec une probabilité $1-\alpha$

$$1-\alpha = \mathrm{Prob}[\theta_1 < \theta < \theta_2]$$

La démarche comprend deux étapes :

- avant le tirage d'un échantillon de taille n, un estimateur $\hat{\theta}$ a été choisi et la loi de probabilité de $\hat{\theta}$ permet de construire un intervalle aléatoire noté $\left|g_1(\hat{\theta}), g_2(\hat{\theta})\right|$ susceptible de contenir la valeur du paramètre θ avec une probabilité 1-α fixée a priori ;
- après tirage, la valeur particulière t de $\hat{\theta}$ calculée à partir des données de l'échantillon permet de déterminer les bornes $g_1(t)$ et $g_2(t)$ de l'intervalle de confiance recherché.

Les paramètres inconnus à estimer seront successivement la moyenne, la variance, la proportion d'une population. Trois autres cas seront considérés : la différence de moyennes, le rapport de variances et la différence de proportions relatives à deux populations.

Rappel du module 1

Paramètres de la population	Estimateurs dans l'échantillon
La moyenne : m	La moyenne dans un échantillon : $\overline{X}$: estimateur $\overline{x}$: valeur calculée
La variance : σ^2	La variance dans un échantillon : S^2 : estimateur s^2 : valeur calculée $\hat{S}^2$: estimateur sans biais $\hat{S}^2 = \frac{n}{n-1} S^2$
La proportion : p	La proportion dans un échantillon : F : estimateur f : valeur calculée

M2Unité 1 : Principe de l'estimation par intervalle de confiance

Soit $\hat{\theta}$ l'estimateur d'un paramètre θ inconnu.

$\hat{\theta}$ est une variable aléatoire dont la loi de probabilité notée $\left(L(\hat{\theta})\right)$ supposée connue dépend de θ. Il est possible de trouver deux valeurs particulieres $t_1(\theta)$ et $t_2(\theta)$ telles que :

$$1-\alpha = \mathrm{Prob}\left[t_1(\theta) < \hat{\theta} < t_2(\theta)\right]$$

S'il est possible de réécrire le système d'inégalités en isolant θ, on peut déterminer un intervalle dont les limites dépendent de $\hat{\theta}$ et tel que :

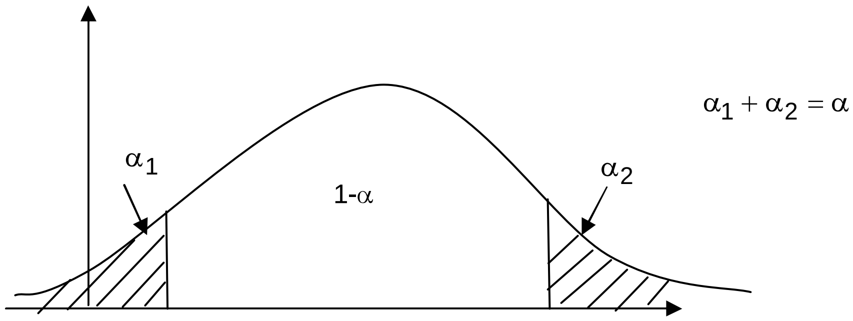
$$1-\alpha = \mathrm{Prob}\left[g_1(\hat{\theta}) < \theta < g_2(\hat{\theta})\right]$$

Ici, l'intervalle qui encadre θ est aléatoire et il possède la propriété de recouvrir la valeur θ dans $1-\alpha$ des cas. La prise en compte d'un échantillon particulier, c'est-à-dire d'une valeur numérique particulière pour $\hat{\theta}$, et donc pour $g_1(\hat{\theta})$ et $g_2(\hat{\theta})$, permet d'obtenir une fourchette qui a de « grandes » chances de « contenir » la valeur inconnue θ si 1-α est élevé.

$$1-\alpha = \mathrm{Prob}\left[g_1(\hat{\theta}) < \theta < g_2(\hat{\theta})\right] = \mathrm{Prob}\left[c_1 < \theta < c_2\right]$$

- $[c_1, c_2]$ ou $\left[g_1(\hat{\theta}), g_2(\hat{\theta})\right]$ est appelé intervalle de confiance,
- c_1, c_2 sont les limites de confiance,
- $1-\alpha$: degré de confiance ou degré de certitude.

Le principe de l'estimation par intervalle de confiance est de proposer un encadrement d'un paramètre inconnu d'une population dont la loi, elle, est connue.



La probabilité α se répartit selon les cas soit à droite d'un certain seuil, soit à gauche, soit à droite et à gauche simultanément.

On parlera :

- d'intervalle bilatéral symétrique si : $\alpha_1 = \alpha_2 = \alpha / 2$
- d'intervalle bilatéral si $\alpha_1 \neq 0$ et $\alpha_2 \neq 0$ avec $\alpha_1 + \alpha_2 = \alpha$
- d'intervalle unilatéral à gauche si $\alpha_2 = 0$
- d'intervalle unilatéral à droite si $\alpha_1 = 0$

Cette démarche appelle trois remarques :

- la probabilté $1-\alpha$ est fixée conventionnellement a priori. Si $1-\alpha = 95\%$, cela signifie que l'intervalle que l'on est susceptible de construire « tombera à côté » de θ (à droite ou à gauche) dans 5% des cas. Si $1-\alpha = 99\%$, ce risque est moins important, mais l'intervalle est plus large ;
- lorsque l'estimateur $\hat{\theta}$ est sans biais, il est naturel de construire un intervalle centré sur l'estimation ponctuelle obtenue pour θ ;

- la détermination de la surface correspondant à la probabilité $(1-\alpha)$ fait intervenir les paramètres permettant de caractériser la distribution de probabilité $\left(L(\hat{\theta})\right)$. Si cette distribution est par exemple normale, deux paramètres interviennent : m et σ.

M2Unité 2 : Estimation par intervalle de confiance de paramètres d'une population

2.1 Estimation par intervalle de confiance de la moyenne d'une population lorsque la variance de la population est connue

Le problème est le suivant : il faut encadrer m (moyenne de la population).

C'est-à-dire on recherche m_1 et m_2 telles que :

$$\exists m_1, m_2 \in \Re \text{ telles que } 1-\alpha = prob[m_1 < m < m_2]$$

On suppose que X obéit à une loi normale $X \equiv N(m, \sigma)$ avec σ connu

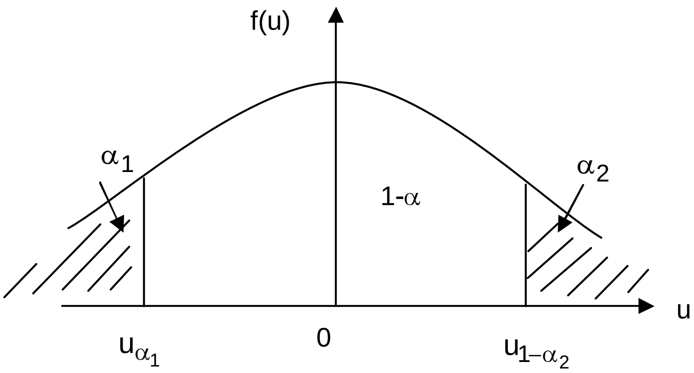
On prélève un échantillon IID de taille n

On sait que $\overline{X} \equiv N\left(m, \sigma/\sqrt{n}\right)$ avec n la taille de l'échantillon (Cf module 1)

Si on forme la variable aléatoire normale centrée réduite $U = \dfrac{\overline{X}-m}{\sigma/\sqrt{n}} \equiv N(0,1)$

Cas d'un intervalle bilatéral

Représentons graphiquement cette loi :



u_{α_1} et $u_{1-\alpha_2}$ sont des valeurs lues dans la table[1] de la loi normale centrée réduite.

A gauche de u_{α_1} nous avons une probabilité α_1 d'où u_{α_1}, à gauche de $u_{1-\alpha_2}$, on a une probabilité cumulée de $1-\alpha_2$ d'où $u_{1-\alpha_2}$

Ecrivons ce que représente graphiquement la probabilité : $1-\alpha$

$$1-\alpha = Prob\left[u_{\alpha_1} < U < u_{1-\alpha_2}\right]$$

Remplaçons U par sa valeur :

[1] Note : on met en indice la valeur de la probabilité correspondant à la fonction de répartition

$$1-\alpha = \mathrm{Prob}\left[u_{\alpha_1} < \frac{\overline{X}-m}{\sigma/\sqrt{n}} < u_{1-\alpha_2}\right]$$

$$1-\alpha = \mathrm{Prob}\left[u_{\alpha_1}\,\sigma/\sqrt{n} < \overline{X}-m < u_{1-\alpha_2}\,\sigma/\sqrt{n}\right]$$

$$= \mathrm{prob}\left[u_{\alpha_1}\,\sigma/\sqrt{n} - \overline{X} < -m < u_{1-\alpha_2}\,\sigma/\sqrt{n} - \overline{X}\right]$$

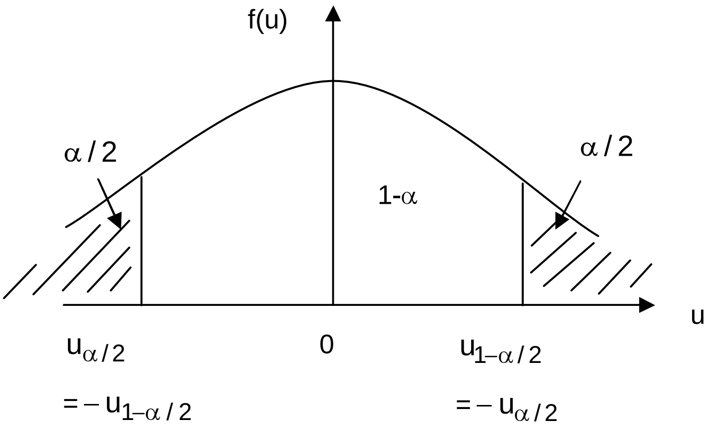
$$= \mathrm{Prob}\left[\underbrace{\overline{X} - u_{1-\alpha_2}\,\sigma/\sqrt{n}}_{m_1} < m < \underbrace{\overline{X} - u_{\alpha_1}\,\sigma/\sqrt{n}}_{m_2}\right]$$

Nous obtenons donc l'encadrement de m ; m est compris entre m_1 et m_2.

Cas particulier : intervalle bilatéral symétrique

Intervalle bilatéral symétrique si : $\alpha_1 = \alpha_2 = \alpha/2$

Graphique 1



$u_{\alpha/2} = -u_{1-\alpha/2}$

$$1-\alpha = \mathrm{Prob}\left[\overline{X} - u_{1-\alpha/2}\,\sigma/\sqrt{n} < m < \overline{X} - u_{\alpha/2}\,\sigma/\sqrt{n}\right]$$

Comme $u_{1-\alpha/2} = -u_{\alpha/2}$ on peut écrire :

$$1-\alpha = \mathrm{Prob}\left[\overline{X} - u_{1-\alpha/2}\,\sigma/\sqrt{n} < m < \overline{X} + u_{1-\alpha/2}\,\sigma/\sqrt{n}\right]$$

Ou encore :

$$1-\alpha = \mathrm{Prob}\left[\overline{X} + u_{\alpha/2}\,\sigma/\sqrt{n} < m < \overline{X} - u_{\alpha/2}\,\sigma/\sqrt{n}\right]$$

Ce qui revient à écrire :

$$1-\alpha = \mathrm{Prob}\left[m \in \left[\overline{X} \pm u_{\alpha/2}\,\sigma/\sqrt{n}\right]\right] \text{ ou } 1-\alpha = \mathrm{Prob}\left[m \in \left[\overline{X} \pm u_{1-\alpha/2}\,\sigma/\sqrt{n}\right]\right]$$

Application :

On cherche un intervalle de degré de confiance bilatéral symétrique à 95% de la moyenne d'une population m. On sait que la variable aléatoire obéit à une loi normale dont l'écart type (connu) est 2 : $X \equiv N(m,2)$. On prélève dans cette population un échantillon de taille n=100.

$$\overline{X} \equiv N\left(m, \sigma/\sqrt{n}\right)$$

$\alpha_1 = \alpha_2 = \alpha/2 = 2{,}5\%$ (intervalle bilatéral symétrique)

$u_{\alpha/2}$ =-1,96 $u_{1-\alpha/2}$ =1,96

$$95\% = \Pr ob\left[-1{,}96 < \frac{\overline{X}-m}{\sigma/\sqrt{n}} < 1{,}96\right]$$
$$= \Pr ob\left[m \in \left[\overline{X} \pm u_{\alpha/2}\ \sigma/\sqrt{n}\right]\right]$$

Supposons que dans l'échantillon la moyenne soit égale à 10 : $\overline{x} = 10$

L'intervalle de confiance calculé est donc : (on le note : $IC_{.95}$)

$$IC_{.95} = [10 \pm 1{,}96\ 2/100] = [9{,}61\,;10{,}39]$$

Il y a une probabilité de 95% que [9,61 ; 10,39] encadre la vraie valeur du paramètre m qui demeure inconnu.

Détermination d'une borne supérieure

On cherche ici m_2 tel que :

$$1-\alpha = \Pr ob[m < m_2] = \Pr ob[-m > -m_2]$$
$$= \Pr ob[\overline{X} - m > \overline{X} - m_2] = \Pr ob\left[\frac{\overline{X}-m}{\sigma/\sqrt{n}} > \frac{\overline{X}-m_2}{\sigma/\sqrt{n}}\right]$$
$$= \Pr ob\left[U > \frac{\overline{X}-m_2}{\sigma/\sqrt{n}}\right]$$

Comme $U \equiv N(0,1)$ on a $1-\alpha = \Pr ob[U > U_\alpha]$

c'est-à-dire :

$$1-\alpha = \Pr ob\left[\frac{\overline{X}-m}{\sigma/\sqrt{n}} > U_\alpha\right]$$
$$= \Pr ob\left[-m > U_\alpha\ \sigma/\sqrt{n} - \overline{X}\right]$$
$$= \Pr ob\left[m < \overline{X} - U_\alpha\ \sigma/\sqrt{n}\right]$$

Pour $\alpha = 5\%$ on a $U_{0,05} = -1{,}6449$

Donc dans note exemple on a la Borne Supérieure suivante :

$$BS = 10 + 1{,}6449\ 2/10 = 10{,}32898 \approx 10{,}33$$

Détermination d'une borne inférieure

Le problème est similaire au précédent, c'est-à-dire que l'on cherche m_1 tel que :

$$1-\alpha = \Pr ob[m > m_1] = \Pr ob[-m < -m_1]$$
$$= \Pr ob\left[\frac{\overline{X}-m}{\sigma/\sqrt{n}} < \frac{\overline{X}-m_1}{\sigma/\sqrt{n}}\right] = \Pr ob\left[\frac{\overline{X}-m}{\sigma/\sqrt{n}} < U_{1-\alpha}\right]$$
$$= \Pr ob\left[-m < U_{1-\alpha}\,\sigma/\sqrt{n} - \overline{X}\right]$$
$$= \Pr ob\left[m > \overline{X} - U_{1-\alpha}\,\sigma/\sqrt{n}\right]$$

Application numérique : $U_{1-\alpha} = U_{0,95} = 1{,}6449$

La borne inférieure est donc : $BI = 10 - 1{,}6449\, {}^{2}\!/_{10} = 9{,}767$

2.2 Estimation par intervalle de confiance de la moyenne d'une population lorsque la variance de la population est inconnue

Le problème est toujours le même : il faut encadrer m.

C'est-à-dire on recherche les m_1 et m_2 telles que :

$\exists m_1, m_2 \in \Re$ telles que $1-\alpha = prob[m_1 < m < m_2]$

$X \equiv N(m,\sigma)$ par hypothèse mais ici σ est inconnu

$\overline{X} \equiv N\left(m, \sigma/\sqrt{n}\right)$ avec n la taille de l'échantillon

Lorsque σ est inconnu, on utilise la loi de Student. (Cf module 1)

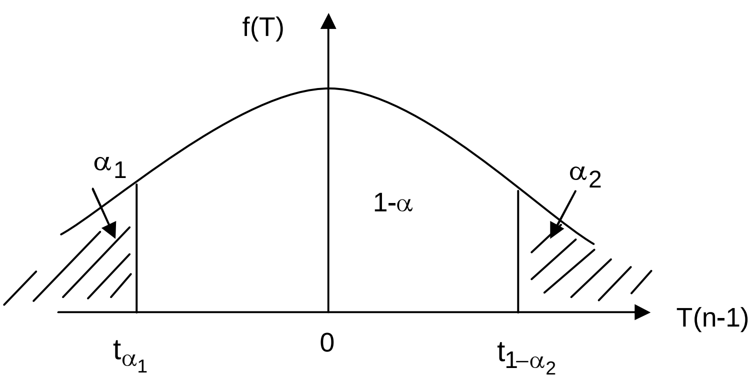
$$\frac{\overline{X}-m}{s/\sqrt{n-1}} \equiv T(n-1)$$

$$\frac{\overline{X}-m}{S/\sqrt{n-1}} = \frac{\overline{X}-m}{\hat{S}/\sqrt{n}} \equiv T(n-1) \text{ car } \hat{S}^2 = \frac{n}{n-1}S^2$$

avec : $\begin{cases} S : \text{estimateur biaisé de } \sigma \\ \hat{S} : \text{estimateur sans biais de } \sigma. \end{cases}$

Cas d'un intervalle bilatéral

Représentons graphiquement cette loi :



t_{α_1} et $t_{1-\alpha_2}$ sont des valeurs lues dans la table de Student. On met ici aussi en indice le seuil de probabilité correspondant à la fonction de répartition.

Ecrivons ce que représente graphiquement la probabilité $1-\alpha$:

$$1-\alpha = \Pr ob\left[t_{\alpha_1} < T(n-1) < t_{1-\alpha_2}\right]$$

Remplaçons T(n-1) par son expression :

$$1-\alpha = \Pr ob\left[t_{\alpha_1} < \frac{\overline{X}-m}{S/\sqrt{n-1}} < t_{1-\alpha_2}\right]$$

$$= \Pr ob\left[t_{\alpha_1} S/\sqrt{n-1} < \overline{X}-m < t_{1-\alpha_2} S/\sqrt{n-1}\right]$$

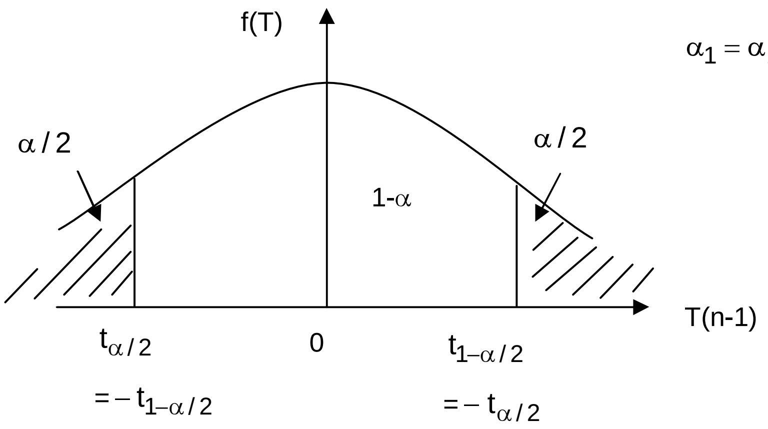
$$= \Pr ob\left[-\overline{X}+t_{\alpha_1} S/\sqrt{n-1} < -m < t_{1-\alpha_2} S/\sqrt{n-1} - \overline{X}\right]$$

$$1-\alpha = \Pr ob\left[\underbrace{\overline{X}-t_{1-\alpha_2} S/\sqrt{n-1}}_{m_1} < m < \underbrace{\overline{X}-t_{\alpha_1} S/\sqrt{n-1}}_{m_2}\right]$$

Nous obtenons donc l'encadrement de m

Cas particulier : intervalle bilatéral symétrique

Graphique 2



$\alpha_1 = \alpha_2 = \alpha/2$

$t_{\alpha/2} = -t_{1-\alpha/2}$

$$1-\alpha = \Pr ob\left[m \in \left[\overline{X} \pm t_{1-\alpha/2} S/\sqrt{n-1}\right]\right]$$

ou encore : $1-\alpha = \Pr ob\left[m \in \left[\overline{X} \pm t_{\alpha/2} S/\sqrt{n-1}\right]\right]$

Intervalle unilatéral à droite

Dans ce cas, $\alpha_1 = 0$ et $\alpha_2 = \alpha$

$$1-\alpha = \text{Prob}\left[\frac{\overline{X}-m}{S/\sqrt{n-1}} < t_{1-\alpha}\right]$$

$$= \text{Prob}\left[m > \underbrace{\overline{X} - t_{1-\alpha}\, S/\sqrt{n-1}}_{m_1}\right]$$

Intervalle unilatéral à gauche

Dans ce cas, $\alpha_1 = \alpha$ et $\alpha_2 = 0$

$$1-\alpha = \text{Prob}\left[\frac{\overline{X}-m}{S/\sqrt{n-1}} > t_{\alpha}\right] = \text{Prob}\left[m < \overline{X} - t_{\alpha}\, S/\sqrt{n-1}\right]$$

$$= \text{Prob}\left[m < \underbrace{\overline{X} + t_{1-\alpha}\, S/\sqrt{n-1}}_{m_2}\right]$$

2.3 Estimation par intervalle de confiance de la variance d'une population

Par hypothèse, X obéit toujours à une loi normale : $X \equiv N(m, \sigma)$

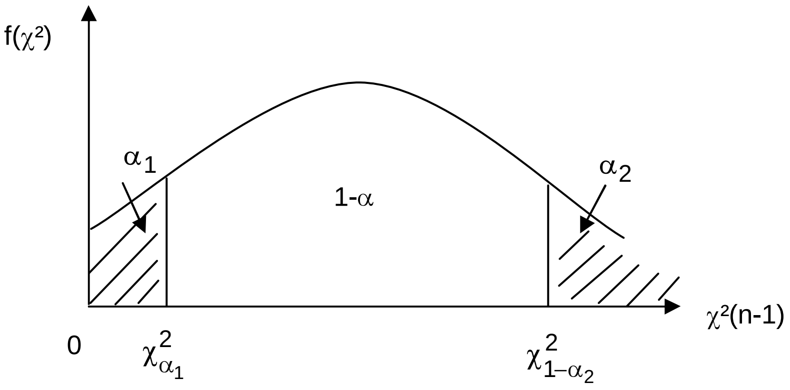
Le problème est le suivant : il faut encadrer σ, l'écart-type de la population qui est inconnu. On recherche donc deux valeurs σ_1^2 et σ_2^2 encadrant σ^2

$$\exists \sigma_1^2, \sigma_2^2 \in \Re \,/\, 1-\alpha = \text{prob}\left[\sigma_1^2 < \sigma^2 < \sigma_2^2\right]$$

On sait que : $\dfrac{nS^2}{\sigma^2} \equiv \chi^2(n-1)$ avec n la taille de l'échantillon. (Cf module 1)

Cas d'un intervalle bilatéral

Représentons graphiquement cette loi :



$\chi^2_{\alpha_1}$ et $\chi^2_{1-\alpha_2}$ sont deux valeurs lues dans la table de la loi du χ^2. On met ici aussi en indice la probabilité correspondant à la fonction de répartition

Ecrivons ce que représente graphiquement la probabilité $1-\alpha$:

$$1-\alpha = \Pr ob\left[\chi^2_{\alpha_1} < \chi^2(n-1) < \chi^2_{1-\alpha_2}\right]$$

Remplaçons $\chi^2(n-1)$ par son expression :

$$1-\alpha = \Pr ob\left[\chi^2_{\alpha_1} < \frac{nS^2}{\sigma^2} < \chi^2_{1-\alpha_2}\right]$$

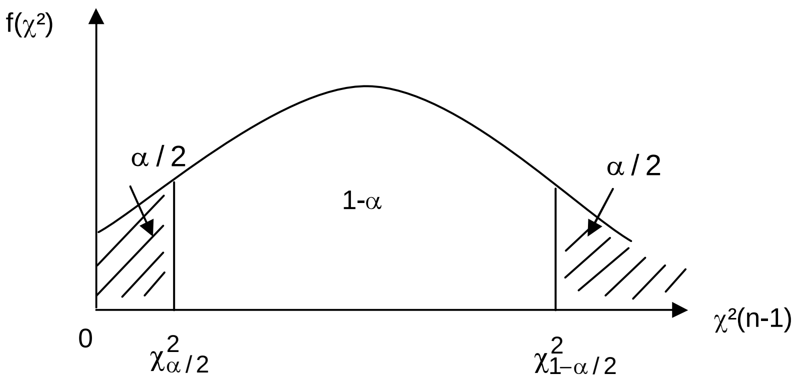
$$= \Pr ob\left[\frac{\chi^2_{\alpha_1}}{nS^2} < \frac{1}{\sigma^2} < \frac{\chi^2_{1-\alpha_2}}{nS^2}\right]$$

$$= \Pr ob\left[\frac{nS^2}{\chi^2_{\alpha_1}} > \sigma^2 > \frac{nS^2}{\chi^2_{1-\alpha_2}}\right]$$

$$1-\alpha = \Pr ob\left[\frac{nS^2}{\chi^2_{1-\alpha_2}} < \sigma^2 < \frac{nS^2}{\chi^2_{\alpha_1}}\right]$$

Nous obtenons ici l'encadrement de σ^2

Cas particulier : intervalle bilatéral symétrique



$$1-\alpha = \Pr ob\left[\frac{nS^2}{\chi^2_{1-\alpha/2}} < \sigma^2 < \frac{nS^2}{\chi^2_{\alpha/2}}\right]$$

ou encore : $1-\alpha = \Pr ob\left[\sigma^2 \in \left[\frac{nS^2}{\chi^2_{1-\alpha/2}};\frac{nS^2}{\chi^2_{\alpha/2}}\right]\right]$

NB : un χ^2 est toujours positif, donc une seule écriture des seuils.

intervalle unilatéral à droite

$$1-\alpha = \Pr ob\left[\frac{nS^2}{\sigma^2} < \chi^2_{1-\alpha}\right] = \Pr ob\left[\frac{nS^2}{\chi^2_{1-\alpha}} < \sigma^2\right]$$

$$= \Pr ob\left[\sigma^2 > \frac{nS^2}{\chi^2_{1-\alpha}}\right]$$

intervalle unilatéral à gauche

$$1-\alpha = \Pr ob\left[\frac{nS^2}{\sigma^2} > \chi^2_{\alpha}\right] = \Pr ob\left[\frac{nS^2}{\chi^2_{\alpha}} > \sigma^2\right]$$

$$= \Pr ob\left[\sigma^2 < \frac{nS^2}{\chi^2_{\alpha}}\right]$$

2.4 Estimation par intervalle de confiance d'une proportion d'une population

Le problème est le suivant : il faut trouver un encadrement de p, c'est-à-dire on recherche deux valeurs p_1 et p_2 telles que

$$\exists p_1, p_2 \,/\, 1-\alpha = \text{prob}[p_1 < p < p_2]$$

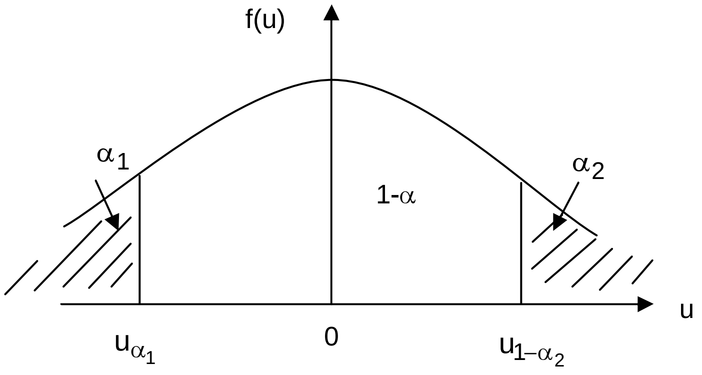
Or sait (Cf module 1) que $F \equiv N\left(p, \sqrt{\frac{pq}{n}}\right)$ avec n la taille de l'échantillon prélevé.

Donc si on forme la variable aléatoire normale centrée réduite, on obtient :

$$\frac{F-p}{\sqrt{\frac{pq}{n}}} \equiv N(0,1)$$

Cas d'un intervalle bilatéral

Représentons graphiquement cette loi :



Ecrivons ce que représente graphiquement la probabilité $1-\alpha$:

$$1-\alpha = \Pr ob\left[u_{\alpha_1} < U < u_{1-\alpha_2}\right]$$

Remplaçons U par sa valeur :

$$1-\alpha = \Pr ob\left[u_{\alpha_1} < \frac{F-p}{\sqrt{\frac{pq}{n}}} < u_{1-\alpha_2}\right]$$

$$= \Pr ob\left[u_{\alpha_1}\sqrt{\frac{pq}{n}} < F-p < u_{1-\alpha_2}\sqrt{\frac{pq}{n}}\right]$$

$$= \Pr ob\left[u_{\alpha_1}\sqrt{\frac{pq}{n}} - F < -p < u_{1-\alpha_2}\sqrt{\frac{pq}{n}} - F\right]$$

$$1-\alpha = \Pr ob\left[\underbrace{F-u_{1-\alpha_2}\sqrt{\frac{pq}{n}}}_{p_1} < p < \underbrace{F-u_{\alpha_1}\sqrt{\frac{pq}{n}}}_{p_2}\right]$$

Nous obtenons donc l'encadrement de p, p compris entre deux valeurs p_1 et p_2.

Le problème ici est que p, la valeur que l'on encadre, se retrouve dans les bornes. Il existe trois méthodes pour donner une valeur aux bornes :

1) Remplacer dans les bornes p et q par leur estimation :

On l'appelle la méthode des estimateurs :

p a pour estimateur sans biais F. $E(\hat{p}) = p$

$p \rightarrow \hat{p} = F \quad q \rightarrow \hat{q} = 1-F$

Remplaçons :

$$1-\alpha = \Pr ob\left[F-u_{1-\alpha_2}\sqrt{\frac{F(1-F)}{n}} < p < F-u_{\alpha_1}\sqrt{\frac{F(1-F)}{n}}\right]$$

Pour un intervalle bilatéral symétrique :

$$1-\alpha = \Pr ob\left[F-u_{1-\alpha/2}\sqrt{\frac{F(1-F)}{n}} < p < F-u_{\alpha/2}\sqrt{\frac{F(1-F)}{n}}\right]$$

$$= \Pr ob\left[p \in \left[F \pm u_{1-\alpha/2}\sqrt{\frac{F(1-F)}{n}}\right]\right]$$

Ou encore : $1-\alpha = \Pr ob\left[p \in \left[F \pm u_{\alpha/2}\sqrt{\frac{F(1-F)}{n}}\right]\right]$

2) Remplacer dans les bornes pq par ¼ : méthode par excès :

Soit p l'estimation de F et q=1-p, le maximum de $pq = \frac{1}{4}$, on considère donc que

$$pq \leq 1/4 \Rightarrow \sqrt{\frac{pq}{n}} \leq \sqrt{\frac{1}{4n}} = \frac{1}{2\sqrt{n}}$$

L'intervalle de confiance bilatéral qui était :

$$1-\alpha = \Pr ob\left[F - u_{1-\alpha_2}\sqrt{\frac{pq}{n}} < p < F - u_{\alpha_1}\sqrt{\frac{pq}{n}}\right]$$

Devient alors :

$$1-\alpha = \Pr ob\left[F - u_{1-\alpha_2}\frac{1}{2\sqrt{n}} < p < F - u_{\alpha_1}\frac{1}{2\sqrt{n}}\right]$$

Cas d'un intervalle bilatéral symétrique :

$$1-\alpha = \Pr ob\left[F - u_{1-\alpha/2}\sqrt{\frac{pq}{n}} < p < F - u_{\alpha/2}\sqrt{\frac{pq}{n}}\right]$$

Comme $u_{1-\alpha/2}\sqrt{\frac{pq}{n}} \leq u_{1-\alpha/2}\frac{1}{2\sqrt{n}}$

L'intervalle bilatéral symétrique devient alors :

$$\Pr ob\left[p \in \left[F \pm u_{1-\alpha/2}\frac{1}{2\sqrt{n}}\right]\right] \geq 1-\alpha$$

3) Méthode de l'ellipse :

Elle s'applique uniquement au cas d'un *intervalle bilatéral symétrique* et ne sera pas détaillée ici. Elle utilise des abaques (courbes dans lesquelles on lit les valeurs des bornes)

2.5 Taille d'un échantillon et précision de l'estimation

2.5.1 Détermination de la taille d'un échantillon en fonction de la précision sur la moyenne

Prenons le cas de l'intervalle bilatéral symétrique[2] de la moyenne lorsque l'écart_type est connu.

On a $$1-\alpha = \Pr ob\left[\underbrace{\overline{X} - u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}}_{m_1} < m < \underbrace{\overline{X} + u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}}_{m_2}\right]$$

Lorsque n augmente, la taille de l'intervalle $[m_1, m_2]$ diminue. $[m_1, m_2]$ dépend à la fois de 1-α et de n.

Réécrivons l'intervalle :

$$1-\alpha = \Pr ob\left[-u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}} < -\overline{X} + m < u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$
$$= \Pr ob\left[\left|\overline{X} - m\right| < u_{1-\alpha/2}\frac{\sigma}{\sqrt{n}}\right]$$

---

[2] On ne peut pas utiliser la cas de l'intervalle bilatéral pour la recherche de la taille d'un échantillon en fonction de la précision. Il faut obligatoirement un intervalle bilatéral symétrique.

On cherche la valeur n telle que la précision sur la moyenne soit égale à C. Le précision sur la moyenne est l'écart qui existe entre la moyenne d'échantillon et la moyenne de la population.

Il existe deux façons de la calculer :

- Précision en valeur absolue

$$1-\alpha = \mathrm{Prob}\left[\left|\overline{X}-m\right| \leq C\right]$$

Le problème est le suivant :

$$\exists n^* \text{ tel que } 1-\alpha = \mathrm{Prob}\left[\left|\overline{X}-m\right| < C\right]$$

On a en outre :

$$1-\alpha = \mathrm{Prob}\left[\left|\overline{X}-m\right| < u_{1-\alpha/2}\frac{\sigma}{\sqrt{n^*}}\right]$$

D'où $\left|\overline{X}-m\right| \leq u_{1-\alpha/2}\frac{\sigma}{\sqrt{n^*}} \leq C \Rightarrow \sqrt{n^*} \geq u_{1-\alpha/2}\frac{\sigma}{C}$

$$\boxed{\Leftrightarrow n^* \geq u_{1-\alpha/2}^2\frac{\sigma^2}{C^2}}$$

Pour déterminer n*, il faut connaître $u_{1-\alpha/2}$, C et σ^2. Si σ^2 n'est pas connu, on passe à la loi de Student (on prend s^2 et $t_{1-\alpha/2}$).

- Précision en valeur relative

Le problème est le suivant :

$$\exists n^* \text{tel que } 1-\alpha = \mathrm{Prob}\left[\frac{\left|\overline{X}-m\right|}{m} \leq C\right]$$

$$\Leftrightarrow 1-\alpha = \mathrm{Prob}\left[\left|\overline{X}-m\right| < Cm\right]$$

De plus on dispose de : $1-\alpha = \mathrm{Prob}\left[\left|\overline{X}-m\right| < u_{1-\alpha/2}\frac{\sigma}{\sqrt{n^*}}\right]$

D'où : $\left|\overline{X}-m\right| \leq u_{1-\alpha/2}\frac{\sigma}{\sqrt{n^*}} \leq Cm \Rightarrow \sqrt{n^*} \geq u_{1-\alpha/2}\frac{\sigma}{\sqrt{n^*}}$

$$\boxed{\Leftrightarrow n^* \geq u_{1-\alpha/2}^2\left(\frac{\sigma}{n^*}\right)^2\frac{1}{C^2}}$$

Si $\frac{\sigma^2}{m^2}$ est inconnu, on tire un échantillon auxiliaire et on calcule $\overline{x}$ et s^2.

2.5.2 Détermination de la taille d'un échantillon en fonction de la précision sur une proportion

Le principe est le même que précédement

- Précision en valeur absolue

On cherche n* tel que : $\Pr ob\left[|F-p| < C\right] = 1-\alpha$

n est la taille nécessaire pour que l'erreur sur l'estimation soit inférieure à C, avec une probabilité égale à 1-α. De plus l'intervalle de confiance de la proportion est égal à :

$$1-\alpha = \Pr ob\left[-u_{1-\alpha/2}\sqrt{\frac{pq}{n^*}} < p-F < u_{1-\alpha/2}\sqrt{\frac{pq}{n^*}}\right]$$

$$1-\alpha = \Pr ob\left[|p-F| \leq u_{1-\alpha/2}\sqrt{\frac{pq}{n^*}}\right]$$

D'où : $|p-F| \leq u_{1-\alpha/2}\sqrt{\frac{pq}{n^*}} \leq C$

$\Leftrightarrow |p-F| \leq u_{1-\alpha/2}\dfrac{1}{2\sqrt{n^*}} \leq C$ si on utilise la méthode par excès.

$$\boxed{\Rightarrow n^* \geq u^2_{1-\alpha/2}\frac{1}{4C^2}}$$

- Précision en valeur relative

$$\exists n^* \text{ tel que } 1-\alpha = \Pr ob\left[\frac{|F-p|}{p} < C\right]$$

$$1-\alpha = \Pr ob\left[F-u_{1-\alpha/2}\sqrt{\frac{pq}{n^*}} < p < F+u_{1-\alpha/2}\sqrt{\frac{pq}{n^*}}\right]$$

On a donc :

$$\begin{cases} 1-\alpha = \Pr ob\left[|p-F| \leq u_{1-\alpha/2}\sqrt{\frac{pq}{n^*}}\right] \\ 1-\alpha = \Pr ob\left[|p-F| \leq Cp\right] \end{cases}$$

D'où : $|p-F| \leq u_{1-\alpha/2}\sqrt{\frac{pq}{n^*}} \leq Cp$

$$\sqrt{n^*} \geq u_{1-\alpha/2}\frac{\sqrt{pq}}{Cp}$$

$\boxed{n^* \geq u^2_{1-\alpha/2}\dfrac{pq}{C^2p^2} = u^2_{1-\alpha/2}\dfrac{1/4}{C^2p^2}}$ Si on utilise la méthode par excès.

M2Unité 3 : Estimation par intervalle de confiance de la différence et du rapport de deux paramètres de deux populations tirées d'une loi normale : cas de deux populations, deux échantillons

Hypothèse : nous avons deux variables aléatoires X_1 et X_2 obéissant à deux lois normales. On prélève dans chacune des populations deux échantillons IID.

$X_1 \equiv N(m_1, \sigma_1)$ taille de l'échantillon prélevé : n_1

$X_2 \equiv N(m_2, \sigma_2)$ taille de l'échantillon prélevé : n_2

3.1 Intervalle de confiance de la différence de deux moyennes lorsque les variances des deux populations sont connues

On cherche deux valeurs c et d telles que : $1-\alpha = \Pr ob[c < m_1 - m_2 < d]$

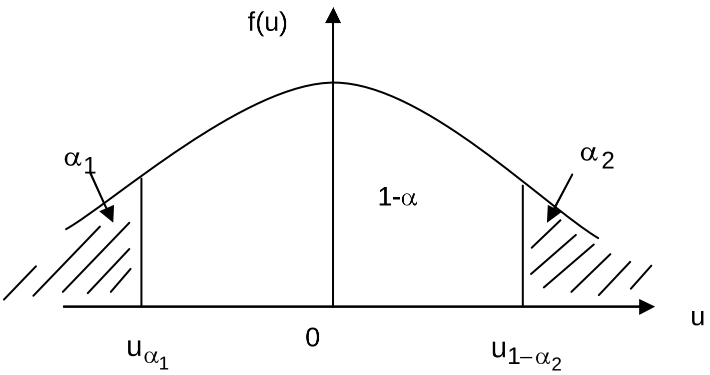
On a vu que : $(\overline{X}_1 - \overline{X}_2) \equiv N\left(m_1 - m_2; \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right)$ σ_i connus. (Cf module 1)

D'où la variable aléatoire normale centrée réduite :

$$U = \frac{(\overline{X}_1 - \overline{X}_2) - (m_1 - m_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \equiv N(0,1)$$

Comme dans le paragraphe précédent, on écrit ce que représente graphiquement la probabilité $1-\alpha$

Plaçons-nous dans le *cas d'un intervalle bilatéral* : $\alpha_1 + \alpha_2 = \alpha$



$$1-\alpha = \Pr ob\left[u_{\alpha_1} < U < u_{1-\alpha_2}\right]$$

Remplaçons U par sa valeur :

$$1-\alpha = \Pr ob\left[u_{\alpha_1} < \frac{(\overline{X}_1 - \overline{X}_2) - (m_1 - m_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} < u_{1-\alpha_2}\right]$$

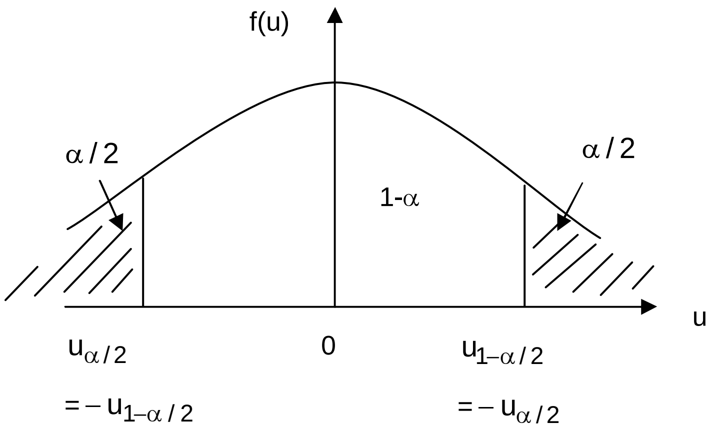
$$= \Pr ob\left[u_{\alpha_1}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < (\overline{X}_1 - \overline{X}_2) - (m_1 - m_2) < u_{1-\alpha_2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$$

$$= \Pr ob\left[u_{\alpha_1}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} - (\overline{X}_1 - \overline{X}_2) < -(m_1 - m_2) < u_{1-\alpha_2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} - (\overline{X}_1 - \overline{X}_2)\right]$$

$$1-\alpha = \Pr ob\left[\left(\overline{X}_1 - \overline{X}_2\right) - u_{1-\alpha_2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < m_1 - m_2 < \left(\overline{X}_1 - \overline{X}_2\right) - u_{\alpha_1}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]$$

Cas d'un intervalle bilatéral symétrique : $\alpha_1 = \alpha_2 = \alpha/2$

Graphique 1



$u_{\alpha/2} = -u_{1-\alpha/2}$

$$1-\alpha = \Pr ob\left[(m_1 - m_2) \in \left[\left(\overline{X}_1 - \overline{X}_2\right) \pm u_{1-\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}\right]\right]$$

avec $\pm u_{1-\alpha/2} = \pm u_{\alpha/2}$ et $u_{1-\alpha/2} = -u_{\alpha/2}$

3.2 Intervalle de confiance de la différence de deux moyennes lorsque les variances des deux populations sont inconnues

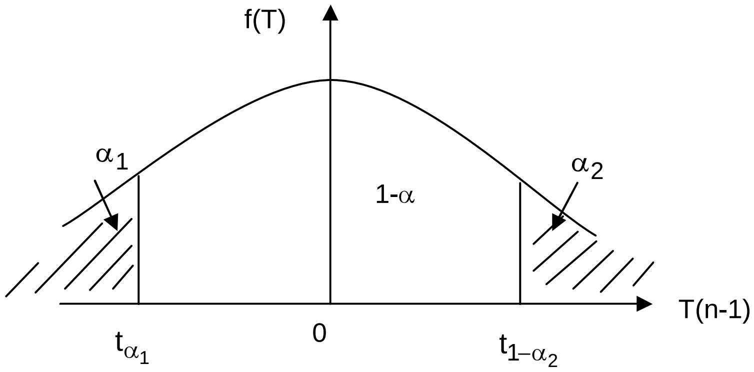
On cherche toujours c et d telles que $1-\alpha = \Pr ob[c < m_1 - m_2 < d]$

$\dfrac{\left(\overline{X}_1 - \overline{X}_2\right) - (m_1 - m_2)}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}} \equiv N(0{,}1)$ avec σ_1 et σ_2 inconnus ici. On utilise donc la loi de Student

Par hypothèse : $\sigma_1 = \sigma_2 = \sigma$

On a vu que (Cf Module 1)

$$\frac{\left(\overline{X}_1 - \overline{X}_2\right) - (m_1 - m_2)}{\sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}\sqrt{\dfrac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}}} \equiv T(n_1 + n_2 - 2)$$



Ecrivons ce que représente graphiquement la probabilité $1-\alpha$ et plaçons-nous dans le *cas d'un intervalle bilatéral* : $\alpha_1 + \alpha_2 = \alpha$

$$1-\alpha = \mathrm{Prob}\left[t_{\alpha_1} < T(n_1+n_2-2) < t_{1-\alpha_2}\right]$$

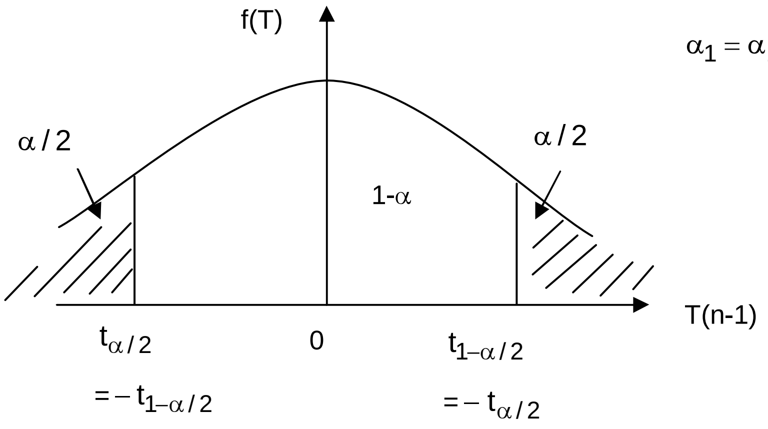
$$1-\alpha = \mathrm{Prob}\left[t_{\alpha_1} < \frac{\left(\overline{X}_1-\overline{X}_2\right)-\left(m_1-m_2\right)}{\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}}} < t_{1-\alpha_2}\right]$$

$$1-\alpha = \mathrm{Prob}\left[t_{\alpha_1}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}} < \left(\overline{X}_1-\overline{X}_2\right)-\left(m_1-m_2\right) < t_{1-\alpha_2}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}}\right]$$

$$1-\alpha = \mathrm{Prob}\left[t_{\alpha_1}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}}-\left(\overline{X}_1-\overline{X}_2\right) < -\left(m_1-m_2\right) < t_{1-\alpha_2}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}}-\left(\overline{X}_1-\overline{X}_2\right)\right]$$

$$1-\alpha = \mathrm{Prob}\left[\left(\overline{X}_1-\overline{X}_2\right)-t_{1-\alpha_2}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}} < \left(m_1-m_2\right) < \left(\overline{X}_1-\overline{X}_2\right)-t_{\alpha_1}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{\frac{n_1S_1^2+n_2S_2^2}{n_1+n_2-2}}\right]$$

Cas d'un intervalle bilatéral symétrique : $\alpha_1 = \alpha_2 = \alpha/2$



$\alpha_1 = \alpha_2 = \alpha / 2$

$t_{\alpha/2} = -t_{1-\alpha/2}$

$$1-\alpha = \Pr ob\left[(m_1 - m_2) \in \left[(\overline{X}_1 - \overline{X}_2) \pm t_{1-\alpha/2}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\sqrt{\frac{n_1 S_1^2 + n_2 S_2^2}{n_1 + n_2 - 2}}\right]\right]$$

avec $t_{\alpha/2} = -t_{1-\alpha/2}$ et $t_{1-\alpha/2} = -t_{\alpha/2}$

3.3 Intervalle de confiance du rapport des variances de deux populations normales

On cherche b et c telles que : $1-\alpha = \Pr ob\left[b < \frac{\sigma_1^2}{\sigma_2^2} < c\right]$ ou $1-\alpha = \Pr ob\left[b < \frac{\sigma_2^2}{\sigma_1^2} < c\right]$

Rappels :

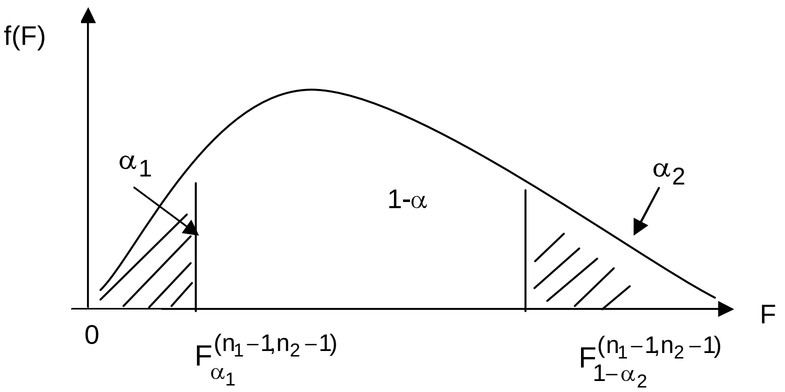
$X_1 \equiv N(m_1, \sigma_1)$ taille de l'échantillon n_1 $\frac{n_1 S_1^2}{\sigma_1^2} \equiv \chi^2(n_1 - 1)$ (Cf Module 1)

$X_2 \equiv N(m_2, \sigma_2)$ taille de l'échantillon n_2 $\frac{n_2 S_2^2}{\sigma_2^2} \equiv \chi^2(n_2 - 1)$ (Cf Module 1)

$$F(n_1, n_2) \equiv \frac{\chi^2(n_1) / n_1}{\chi^2(n_2) / n_2}$$ Indépendants (Cf Module 1)

$$F(n_1 - 1; n_2 - 1) = \frac{n_1 S_1^2}{(n_1 - 1)} \cdot \frac{\sigma_2^2}{\sigma_1^2} \cdot \frac{(n_2 - 1)}{n_2 S_2^2}$$

Comme $\hat{S}^2 = \frac{n}{n-1} S^2 \Rightarrow F(n_1 - 1; n_2 - 1) = \frac{\hat{S}_1^2}{\hat{S}_2^2} \cdot \frac{\sigma_2^2}{\sigma_1^2}$



Plaçons nous dans le *cas d'un intervalle bilatéral* $\alpha_1+\alpha_2=\alpha$ et écrivons ce que représente graphiquement la probabilité $1-\alpha$:

$$1-\alpha=\operatorname{Prob}\left[F_{\alpha_1}<F(n_1-1;n_2-1)<F_{1-\alpha_2}\right]$$

$$1-\alpha=\operatorname{Prob}\left[F_{\alpha_1}<\frac{\hat{S}_1^2}{\hat{S}_2^2}\frac{\sigma_2^2}{\sigma_1^2}<F_{1-\alpha_2}\right]$$

$$1-\alpha=\operatorname{Prob}\left[F_{\alpha_1}\frac{\hat{S}_2^2}{\hat{S}_1^2}<\frac{\sigma_2^2}{\sigma_1^2}<F_{1-\alpha_2}\frac{\hat{S}_2^2}{\hat{S}_1^2}\right]$$ si on encadre $\frac{\sigma_2^2}{\sigma_1^2}$

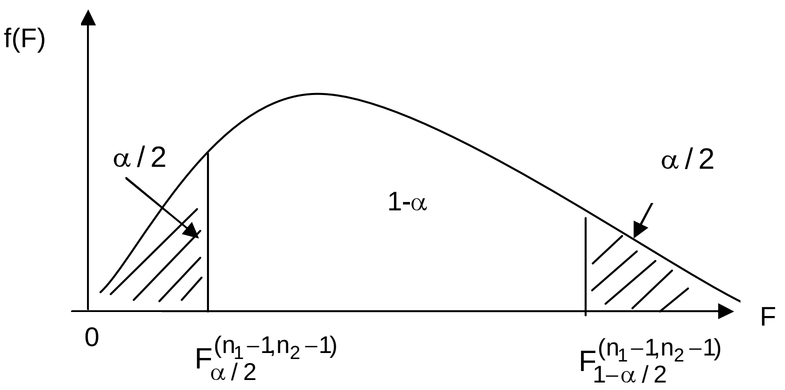
ou $$1-\alpha=\operatorname{Prob}\left[\frac{\hat{S}_1^2}{\hat{S}_2^2}\frac{1}{F_{1-\alpha_2}^{(n_1-1,n_2-1)}}<\frac{\sigma_1^2}{\sigma_2^2}<\frac{\hat{S}_1^2}{\hat{S}_2^2}\frac{1}{F_{\alpha_1}^{(n_1-1,n_2-1)}}\right]$$ si on encadre $\frac{\sigma_1^2}{\sigma_2^2}$

Rappel : $F_{\alpha_1}(n_1,n_2)=\dfrac{1}{F_{1-\alpha_1}(n_2,n_1)}$

Autre méthode : on forme

$F(n_2-1,n_1-1)=\dfrac{\hat{S}_2^2}{\hat{S}_1^2}\cdot\dfrac{\sigma_1^2}{\sigma_2^2}$ et on encadre directement $\dfrac{\sigma_1^2}{\sigma_2^2}$

Cas d'un intervalle bilatéral symétrique : $\alpha_1=\alpha_2=\alpha/2$



$$1-\alpha = \text{Prob}\left[F_{\alpha/2}\frac{\hat{S}_2^2}{\hat{S}_1^2} < \frac{\sigma_2^2}{\sigma_1^2} < F_{1-\alpha/2}\frac{\hat{S}_2^2}{\hat{S}_1^2}\right]$$

ou $$1-\alpha = \text{Prob}\left[\frac{\hat{S}_1^2}{\hat{S}_2^2}\frac{1}{F_{1-\alpha/2}^{(n_1-1,n_2-1)}} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{\hat{S}_1^2}{\hat{S}_2^2}\frac{1}{F_{\alpha/2}^{(n_1-1,n_2-1)}}\right]$$

3.4 Estimation par intervalle de confiance de la différence de deux proportions

On cherche à déterminer deux valeurs c et d telles que : $1-\alpha = \text{Prob}[c < p_1 - p_2 < d]$

Nous avons deux populations dans lesquelles ont peut définir deux modalités A et $\overline{A}$ ayant pour probabilités respectives p_1 et q_1 dans la population 1 (p_2 et q_2 dans la population 2)

	Population 1	Population 2
Modalité A	p_1	p_2
Modalité $\overline{A}$	$1-p_1 = q_1$	$1-p_2 = q_2$
	$p_1 + q_1 = 1$	$p_2 + q_2 = 1$

p_1 et p_2 inconnus.

On tire un échantillon dans chacune des populations : on définit alors deux variables aléatoires F_1 et F_2.

$$\hat{p}_1 = F_1 \quad F_1 = \frac{Y_1}{n_1}$$

$$\hat{p}_2 = F_2 \quad F_2 = \frac{Y_2}{n_2}$$

<u>Rappels du module 1 :</u>

$$Y_1 \equiv B(n_1,p_1) \to N\left(n_1p_1,\sqrt{n_1p_1q_1}\right) \quad F_1 \to N\left(p_1,\sqrt{\frac{p_1q_1}{n_1}}\right)$$

$$Y_2 \equiv B(n_2,p_2) \to N\left(n_2p_2,\sqrt{n_2p_2q_2}\right) \quad F_2 \to N\left(p_2,\sqrt{\frac{p_2q_2}{n_2}}\right)$$

$$F_1 - F_2 \to N\left(p_1 - p_2;\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}\right)$$

Ecrivons ce que représente la probabilité $1-\alpha$ et plaçons nous dans le cadre d'un intervalle bilatéral $\alpha_1 + \alpha_2 = \alpha$:

$$1-\alpha = \text{Prob}\left[u_{\alpha_1} < U < u_{1-\alpha_2}\right] \quad U \equiv N(0,1)$$

$$1-\alpha = \text{Prob}\left[u_{\alpha_1} < \frac{(F_1-F_2)-(p_1-p_2)}{\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}} < u_{1-\alpha_2}\right]$$

$$1-\alpha=\Pr ob\left[u_{\alpha_1}\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}<(F_1-F_2)-(p_1-p_2)<u_{1-\alpha_2}\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}\right]$$

$$1-\alpha=\Pr ob\left[u_{\alpha_1}\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}-(F_1-F_2)<-(p_1-p_2)<u_{1-\alpha_2}\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}-(F_1-F_2)\right]$$

$$1-\alpha=\Pr ob\left[(F_1-F_2)-u_{1-\alpha_2}\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}<p_1-p_2<(F_1-F_2)-u_{\alpha_1}\sqrt{\frac{p_1q_1}{n_1}+\frac{p_2q_2}{n_2}}-\right]$$

Il est possible d'utiliser les méthodes vues dans le paragraphe 2.4.

1) Remplacer dans les bornes p et q par leur estimation :

$p_1 \rightarrow \hat{p}_1 = F_1$

$p_2 \rightarrow \hat{p}_2 = F_2$

$$1-\alpha=\Pr ob\left[(F_1-F_2)-u_{1-\alpha_2}\sqrt{\frac{F_1(1-F_1)}{n_1}+\frac{F_2(1-F_2)}{n_2}}<p_1-p_2<(F_1-F_2)-u_{\alpha_1}\sqrt{\frac{F_1(1-F_1)}{n_1}+\frac{F_2(1-F_2)}{n_2}}\right]$$

2) Remplacer dans les bornes pq par ¼ : méthode par excès :

$$1-\alpha=\Pr ob\left[(F_1-F_2)-u_{1-\alpha_2}\frac{1}{2}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}<p_1-p_2<(F_1-F_2)-u_{\alpha_1}\frac{1}{2}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]$$

Pour un intervalle bilatéral symétrique :

$$1-\alpha=\Pr ob\left[(p_1-p_2)\in\left[(F_1-F_2)\pm u_{1-\alpha/2}\sqrt{\frac{F_1(1-F_1)}{n_1}+\frac{F_2(1-F_2)}{n_2}}\right]\right] \text{ : première méthode}$$

$$1-\alpha=\Pr ob\left[(p_1-p_2)\in\left[(F_1-F_2)\pm u_{\alpha/2}\frac{1}{2}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}\right]\right] \text{ : deuxième méthode}$$